Supporting Cancer prevention strategies using Geospatial Analysis on HRSA data

CP 6950- GIS Capstone Project; Spring 2016

Arthi Rao, Doctoral Candidate
School of City and Regional Planning, Georgia Institute of Technology

Introduction

The Affordable Care Act of 2010 greatly expanded the role of Federally Qualified Health Centers (FQHCs) in providing healthcare services to underserved communities. The American Cancer Society (ACS) and other public health organizations consider FQHCs as key collaborators in a national effort to promote disease prevention. ACS leads cancer prevention initiatives such as colorectal (80% by 2018) and breast cancer screening, and recently is working with the CDC to increase HPV vaccination uptake.

FQHCs are required to report on a number of measures regarding the operation and performance of health centers. These include clinical, operational and financial data as well as aggregate patient characteristics including health outcomes, services and residential zip codes. Among these measures are rates for colorectal cancer screening, breast cancer screening and cervical cancer screening (pap smears).

Geographic coverage of service areas for health providers and healthcare utilization is available through organizations such as HRSA, Dartmouth Atlas and others. These types of spatial data are routinely used in health policy research. However, the spatial coverage of safety net providers such as FQHCs has not been studied. The goal of this project was to develop a methodology to generate and visualize FQHC service areas. The project also demonstrates how geospatial analysis of HRSA and Census data can identify factors associated with disparities in cancer screening rates and redirect available resources strategically towards local cancer prevention efforts and other targeted interventions.

Both contiguity and non-contiguity based methods were compared and “core” service areas (SAs) were generated using Python scripts (PHASE I). The methodology is validated by examining both spatial and temporal trends within the geographical distribution of service areas for 2012, 2013 and 2014 data. Core service areas are layered with socioeconomic profiles, urban/rural classifications, cancer screening rates and other data from publicly available sources, such as the Census and BRFSS. Screening rates for colorectal cancer as reported by FQHCs are also mapped to demonstrate the utility of this methodology. The analysis identified "Communities of Focus" where there is high co-occurrence of risk factors (PHASE II).

The advanced spatial analysis and modeling were effective in identifying areas of low screening rates and associated socioeconomic determinants. This integrated information will be used to evaluate performance of ACS cancer screening initiatives and to support better targeting and improved resource allocation in communities. Data, methods, and a collaborative roadmap for utilizing this research within the organization are elucidated in this paper.

A brief history of FQHCs

Federally Qualified Health Centers were established in 1965 during the Lyndon Johnson administration as part of the War on Poverty program. Neighborhood (Community) Health centers were commissioned by the Office of Economic Opportunity to provide poor and medically underserved communities with access to health and social services (safety-net provideers). These health centers were formally sanctioned and funded under Section 330 of the Public Health Service Act (PHSA). The grant funding program is administered by the Health Resources and Services Administration (HRSA), Bureau of Primary Health Care within the U.S. Department of health and Human Services (HHS). FQHCs receive grant funds from HHS to cover the cost of providing care to uninsured patients. They also bill Medicaid, Medicare and private insurers for services provided to insured patients. In order to receive section 330 grant funds, an FQHC must meet the following requirements (Taylor, 2012):

- Must be located in a federally designated medically underserved area (MUA) or serve a federally designated medically underserved population (MUP)
- Have nonprofit or public status
- Provide comprehensive primary health care services and referrals
- Provide services such as case management, translation and transportation that facilitate access to care
- Have a governing board where majority of the members are patients of the health center
- Provide services to everyone regardless of their ability to pay and have a sliding fee schedule that adjusts to the family income.

FQHCs and Cancer Prevention

FQHCs provide comprehensive primary and preventive care to patients regardless of their ability to pay for services. The Affordable Care Act (ACA) greatly expanded the role of FQHCs by increasing infrastructural support as well as extension of Medicaid coverage to low-income adults, irrespective of their family status. The ACA is predicted to create an influx of newly-insured, low income patients into FQHCs. These patients are at higher risk for cancer due to their socioeconomic status. In addition to disproportionate cancer morbidity and mortality, the uninsured and underinsured also get much lower levels of preventive services including screening for breast, colorectal and cervical cancer compared to the general population. However, several measures including increased funding, patient-centered medical home transformation, electronic health record implementation and increased emphasis on screening from HRSA is anticipated to increase screening levels within FQHCs (Allen et al., 2013; Daly et al, 2015; Martinez-Gutierrez et al., 2013; Taplin et al., 2008). Cancer-focused organizations such as the ACS are seeking to collaborate with FQHCs to increase cancer screening and other prevention strategies.

Significance of Study

Geographic datasets such as Hospital Referral Regions (HRRs), Hospital service areas (HSAs), Pediatric surgical areas (PSAs) and Primary care service areas (PCSAs) are available through a joint effort between HRSA and The Dartmouth Institute for Health Policy and Clinical Practice. These datasets are freely downloadable at the Dartmouth Atlas for Healthcare website. A key limitation to this data is that it only captures Medicare related patient utilization patterns. These datasets are also developed by aggregating different spatial units and capture related but different aspects of the healthcare system.

To date, the spatial distribution and coverage of FQHC utilization patterns (safety-net services) have not been studied at a national scale. If interventions for improving screening are to be effective, it would be useful to visualize the geographical patterns of FQHC utilization and understand their socioeconomic/demographic composition. Organizations can then strategically plan their interventions by identifying “Communities of Focus” or areas of low screening rates where risk factors coalesce.

The American Cancer Society plays a primary role in prevention initiatives such as the 80% by 2018 convened through the National Colorectal Cancer Roundtable (NCCRT) jointly founded by the CDC and ACS and is a coalition between several agencies. The aim is to increase colorectal cancer screening rates to 80% by the year 2018. Breast and cervical cancer screening and tobacco cessation are other important areas of focus. The ACS also funds FQHCs to improve their screening programs. An analysis of FQHC service areas and understanding their constituent communities are foreseen to help all of these initiatives. This paper primarily focuses on an analysis of colorectal cancer screening.

Healthcare Service Areas

Service areas, also referred to as catchment areas, are geographic areas that capture the patient population utilizing a particular hospital or other healthcare provider. These areas usually represent patient travel patterns and demonstrate variations in patterns of utilization (Klauss et al, 2005). Several variations on the definitions and methodologies used to construct these service areas exist based on their intended purpose. Some commonly used typologies are described in table 1.

Table 1. Typologies of commonly available healthcare service areas

Service Area	Definition	Methodology
Hospital service areas (HSAs)	Polygons representing local health care markets for hospital care. An HSA is a collection of ZIP codes whose residents receive most of their hospitalizations from the hospitals in that area.	HSAs were defined by assigning ZIP codes to the hospital area where the greatest proportion of their Medicare residents were hospitalized. Minor adjustments were made to ensure geographic contiguity. This process resulted in 3,436 HSAs. When these regions were created in the early 1990s, most hospital service areas contained only one hospital. In the intervening years, hospital closures have left some HSAs with no hospital; these HSAs have been maintained as distinct areas in order to preserve the continuity of the database.
Hospital referral regions (HRRs)	Polygons representing regional health care markets for tertiary medical care that generally requires the services of a major referral center. The regions were defined by determining where patients were referred for major cardiovascular surgical procedures and for	Each hospital service area (HSA) was examined to determine where most of its residents went for these services. The result was the aggregation of the 3,436 hospital service areas into 306 HRRs. Each HRR has at least one city where both major cardiovascular surgical procedures and neurosurgery are performed.

	neurosurgery.	
Primary care service areas (PCSAs)	Polygons reflect Medicare patient travel to primary care providers.	Population zip codes were assigned to the provider zip code where the plurality of beneficiaries received care, and these assignments were adjusted to create contiguous zip code groups
Health Service Areas	Polygons representing a single county or cluster of contiguous counties which are relatively self-contained with respect to hospital care	Counties grouped into service areas through agglomerative hierarchical clustering analysis.

Based on the methodology for each service area described in table 1, it can be seen that these readily downloadable datasets focus on Medicare utilization patterns. There is no such data available on FQHCs or other community health providers. Generating and mapping the FQHC service areas would enable visualizing the spatial distribution and coverage of safety-net health services. This exercise would be valuable in and of itself to inform a wide variety of purposes within ACS and within the larger research community engaged in health policy and services research. Within ACS, overlaying socioeconomic and other risk factors over this base layer was foreseen to inform strategic prevention initiatives as well as resource allocation and fund raising.

UDS Data on Federally Qualified Health Centers

The Uniform Data System (UDS) is a reporting requirement for Health Resources and Service Administration (HRSA) grantees, including community health centers, migrant health centers, health care for the homeless grantees, and public housing primary care grantees. FQHCs are required to submit data annually to the Bureau of Primary Health Care (BPHC) through HRSA's Electronic Handbook (EHB) system. The UDS is a core system of information appropriate for reviewing the operation and performance of health centers, identify trends over tie and improve health center performance. UDS data are compared with national data to review differences between the U.S population at large and those individuals and families who rely on the health care safety net for primary care.

The UDS is comprised of 12 tables designed to yield consistent clinical, operational, and financial data:

1. Patient Origin: Patients served reported by ZIP code and by primary third party medical insurance source, if any
2. Table 3A: Patients by age and gender
3. Table 3B: Patients by race, Hispanic/Latino ethnicity, and language barriers
4. Table 4: Patients by income (percent of poverty level) and primary third party medical insurance source; Table 4 also reports the number of "special population" patients receiving services, and managed care member months.
5. Table 5: Full-time equivalent staff by position, visits by provider type, and patients by service type
6. Table 5A: Tenure for selected health center staff
7. Table 6A: Selected diagnoses for medical, mental health, and substance abuse visits; and selected medical and dental services provided
8. Table 6B: Quality of care measures
9. Table 7: Health outcomes measures by race and ethnicity
10. Table 8A: Direct and indirect expenses by cost center
11. Table 9D: Full charges, collections, and allowances by payor type as well as sliding discounts and patient bad debt
12. Table 9E: Non patient-service income

The entire set of 12 tables was obtained through a Freedom of Information Act (FOIA) request. The "Patient Origin" and "Quality of Care measures" tables were utilized in this project. The initial request was made for 2012 data. Subsequently, 2013 and 2014 data were also obtained through the same process. The service area methodology and subsequent analysis documented in this paper was performed on 2012 data.

Tables 2 and 3 provide a snapshot of the data and structure of the variables. Table 2 shows an example of an FQHC identified through the BHCMISID (unique identifier that stays consistent for all years). The “ZipCode”, “ZipCodeType” and “TotalNumberofPatients” are the primary variables of interest. The BHCMISID shares a one-to-many relationship with the other variables.

Table 2. The "Patient Origin" table showing the number of patients from each zipcode seen at an FQHC (the BHCMISID is the unique identifier for an FQHC)

BHCMISID	GrantNumber	ReportingYear	ZipCode	ZipCodeType	TotalNumberofPatients
010040	H80CS00443	2013		Other ZIP Codes	
010040	H80CS00443	2013		Unknown Residence	22
010040	H80CS00443	2013	04001	ZipCode	3
010040	H80CS00443	2013	04002	ZipCode	4
010040	H80CS00443	2013	04008	ZipCode	6
010040	H80CS00443	2013	04011	ZipCode	2
010040	H80CS00443	2013	04037	ZipCode	21
010040	H80CS00443	2013	04049	ZipCode	4
010040	H80CS00443	2013	04062	ZipCode	3
010040	H80CS00443	2013	04084	ZipCode	6
010040	H80CS00443	2013	04102	ZipCode	1
010040	H80CS00443	2013	04104	ZipCode	1
010040	H80CS00443	2013	04210	ZipCode	21
010040	H80CS00443	2013	04220	ZipCode	18
010040	H80CS00443	2013	04236	ZipCode	1
010040	H80CS00443	2013	04240	ZipCode	18
010040	H80CS00443	2013	04258	ZipCode	3
010040	H80CS00443	2013	04259	ZipCode	32
010040	H80CS00443	2013	04266	ZipCode	9

Under ZipCodeType, there are two categories—1) Other ZIP Codes 2) Unknown Residence, that indicate that there are some patients for whom the zipcode is not known or recorded. These categories do not contain a specific spatial reference and cannot be used in the construction of Service Areas. Patients from these categories will not be accounted for in the service areas and geospatial analysis and forms a key limitation of the method. However, a preliminary evaluation was done to assess how much total information would be lost due to their exclusion. There were a total of 1194 unique FQHCs in the 2012 data. 13 FQHCs had less than 80% captured from known zipcodes (table 3).

Table 3. Percentage of patients remaining after eliminating unidentifiable zipcodes

BHCMISID	PatientTot_KnownZip	PatientTot_UnknownOther	Percent
0612210	1990	5050	39.4%
106230	549	1168	47.0%
05E00481	514	894	57.5%
102040	3008	4623	65.1%
04E00489	280	414	67.6%
052710	9589	14167	67.7%
04E00479	334	487	68.6%
1017080	1081	1542	70.1%
102900	1696	2333	72.7%
0450640	1569	2114	74.2%
0928870	4122	5438	75.8%
02E00104	656	862	76.1%
011190	12469	16143	77.2%
05E00238	265	327	81.0%
058480	12654	15584	81.2%
08E00441	1513	1861	81.3%
09E00005	9255	11367	81.4%

The "Quality of Care measures" table (table 4) has numerous measures pertaining to screening snd prevention rates for each FQHC. The colorectal screening rates are used as a demonstration in this study.

Table 4. A sample of the "Quality of Care measures" table where the colorectal cancer screening rates for each FQHC (identified through the BHCMISID) are recorded.

BHCMISID	GrantNumber	NumberofPatients Immunized	%ofPatients Immunized	TotalPatients3rdbirthday measurementyear	#ofAdults51-74withAppropriateScreening forColorectalCancer	TotalofAdults51-74ScreeningforColorectalCancer
010030	H80CS00803	191.714285714286	87.14%	220	1553.65714285714	2862
010040	H80CS00443	1	50%	2	34.7142857142857	270
010060	H80CS00741	216	87.8%	246	976	1982
010070	H80CS00312	148.342857142857	84.29%	176	1475.75714285714	5437
010120	H80CS00001	0	0%	2	249	453
010130	H80CS00002			0	13	278
010150	H80CS00311	23	46.94%	49	144	2791
010160	H80CS00499	37	100%	37	737	1173
010170	H80CS00743	150	88.76%	169	1989	3914
010180	H80CS00003	4	23.53%	17	182.7	609
010220	H80CS00155	192.214285714286	92.86%	207	1603.1	4879
010290	H80CS00810	86	86%	100	74	112
010330	H80CS00601	16	69.57%	23	1223	2017
010340	H80CS00591	22	78.57%	28	2460.04285714286	2823
010380	H80CS00150	5	71.43%	7	392.228571428571	858
010450	H80CS06657	9	90%	10	160	410
010460	H80CS00797	71	55.47%	128	5677	10032
010480	H80CS00308	40	86.96%	46	2372.12857142857	3133
010530	H80CS00396	27	93.1%	29	688	860

Overview of Service Area Methodology (PHASE I)

From a methodological perspective, the primary aim of this project was to create "Core Service Areas" that effectively captured a majority of the FQHC patient population. The core SAs would be a representation of the community surrounding the FQHC. Census tract level socioeconomic and other risk factors were then computed for each SA. Risk factors for colorectal cancer and screening rates were overlaid on the SAs to identify communities of focus for strategic prevention initiatives. The overall process is outlined in fig.1 below.

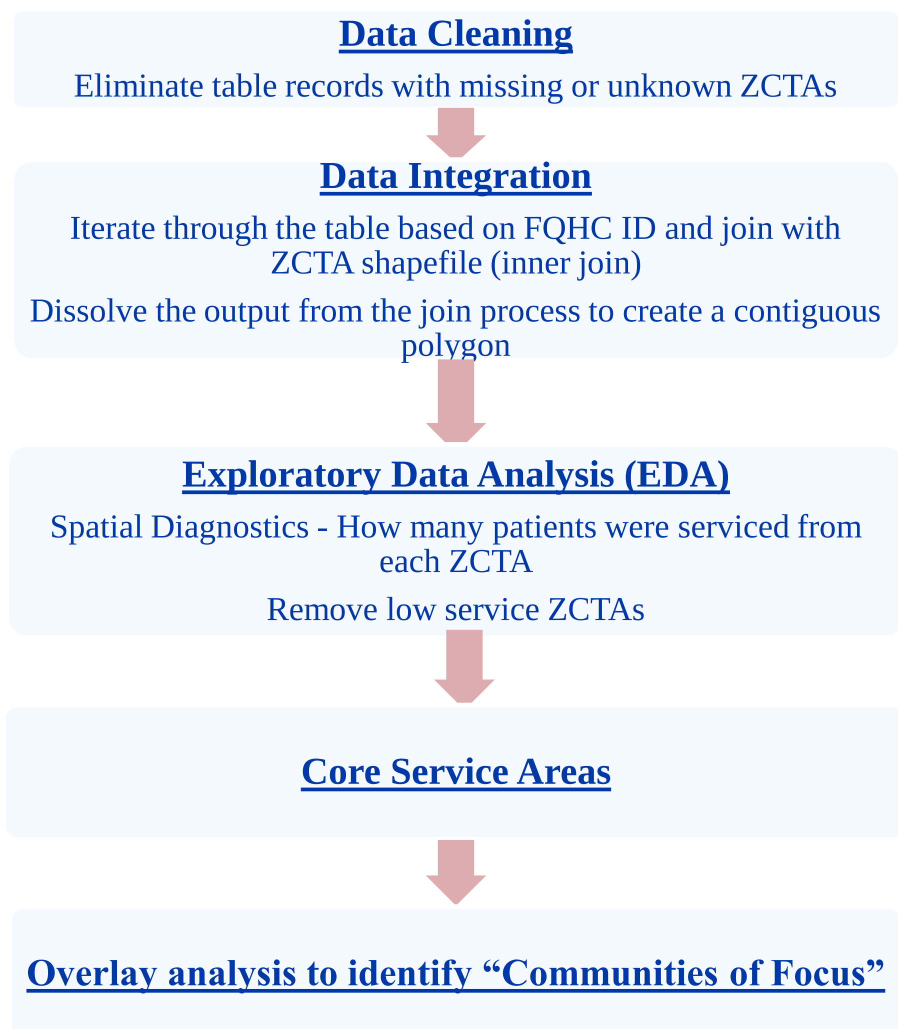


Figure 1. Overview of Methodology

Creating the baseline Service Areas

The cleaned excel file (removal of ZipCodeType with "Other" and "Unknown" categories) was used for generating the first round of Service Areas that included all patients with known zipcodes of origin. Using the census crosswalk files, zipcodes were matched to their respective zctas to prevent loss of patients from zipcodes that were not an exact match with a zcta. Patient counts for each zcta were further aggregated using pivot tables.

Since this was the first attempt at creating the SAs, the decision was made to keep each component of the process separate to retrace steps and be able to inspect every FQHC individually. It was also essential to make each SA available individually for further analysis. The first step was to create separate tables for each unique FQHC and store it within a folder. A free downloadable excel add-in (ASAP utilities) was used to accomplish this task. The algorithm shown in fig.1 was then developed to write and execute the Python scripts.

A discrete polygon shapefile was created for each FQHC and saved in a folder. After the dissolve process, the polygons did not contain any attribute information. However, ideally, each polygon would be able to be identified through the BHCMISID associated with the FQHC that it served. The BHCMISID was extracted from the name of the shapefile and added to an additional field created in the attribute table for each SA shapefile. These steps were also executed using python scripts. The SAs were then mapped to visualize the spatial distribution of SAs including all patients.

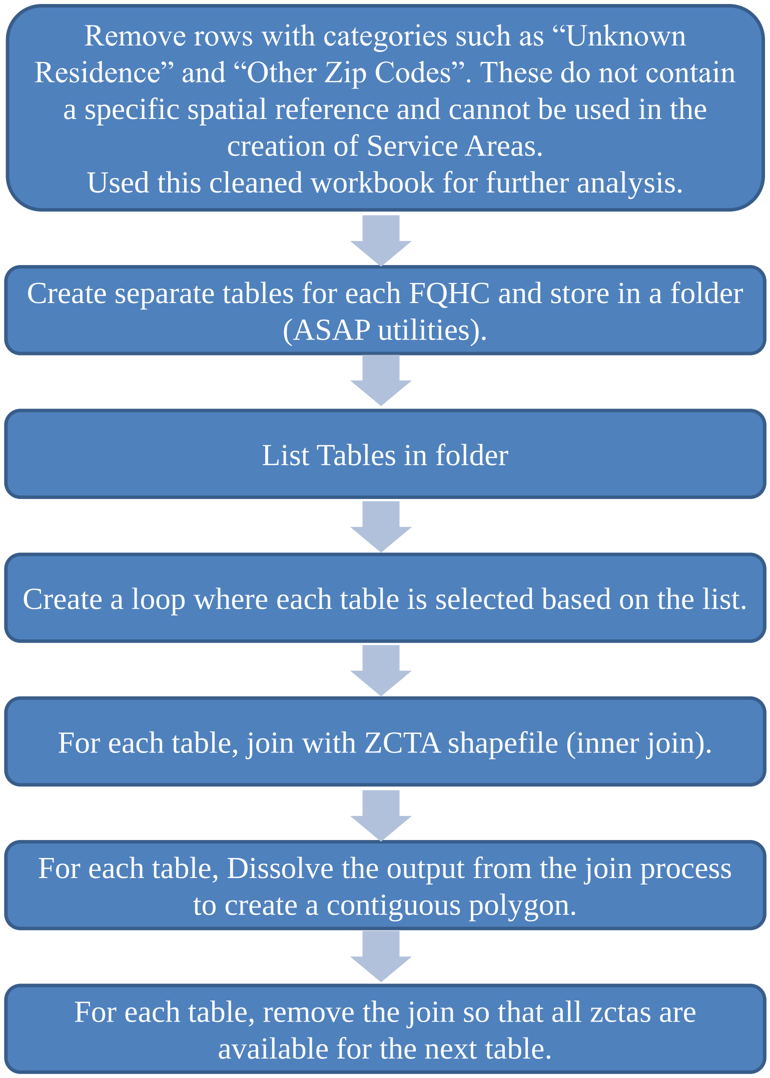


Figure 2. Algorithm used to write Python scripts to generate Service Area polygons.

Exploratory Data Analysis (EDA)

When the Service Areas were mapped from the previous process, the FQHC coverage appears really good geographically. However, this information could be misleading as numerous Service Areas have several distant and proximal zipcodes with just 1 patient. The EDA was executed to examine scenarios where zip codes with small patient populations would be eliminated from the Service Areas and the resulting information that would be lost. The logic of the EDA process is explained in fig.3 below.

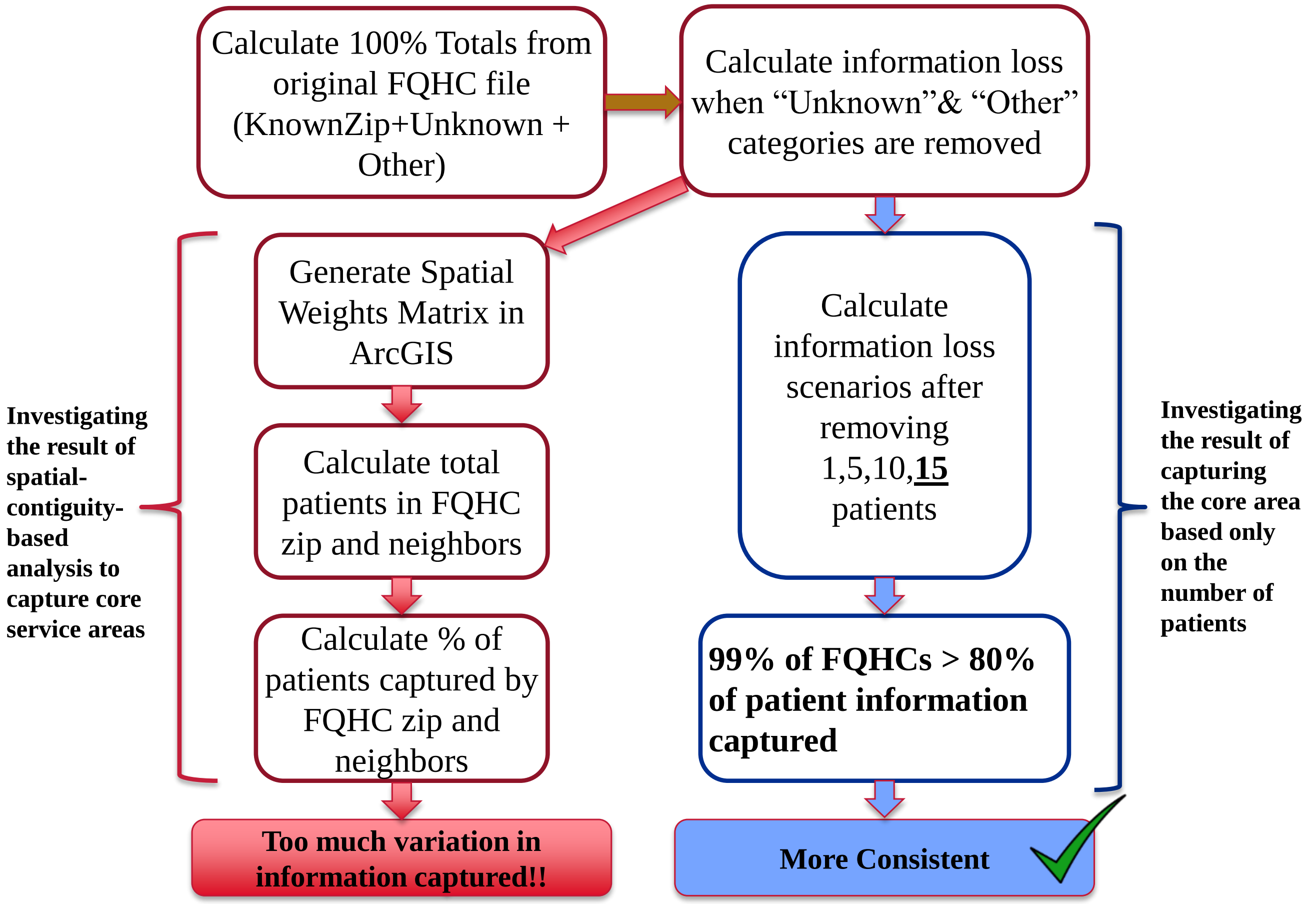


Figure 3. Overview of the EDA process

Both contiguity and non-contiguity based methods were evaluated to inform the methodology. The contiguity based method was built on the assumption that majority of FQHC patients reside in the immediate neighborhood of the FQHC location. This is a commonly used approach to creating service areas as healthcare utilization patterns are assumed to be fairly localized. However, these patterns are regulated through insurance and other mechanisms that determine healthcare access and are less relevant to FQHC access.

The first step in the EDA process was to compare the number of patients captured by aggregating patients from the zipcodes in the immediate neighborhood of each FQHC. This was accomplished by adding up patients from the zctas containing the FQHC as well as those surrounding it as identified from the contiguity matrix generated in ArcGIS (table 5).

Table 5. ArcGIS generates a sparse matrix that only identifies the adjacent zipcodes.

ZIPCODEINT	NID	WEIGHT	BHCMISID_Z	Neighbor_Zi	weight_1
624	601	0.25000000000	00624	00601	1
631	601	0.25000000000	00631	00601	1
641	601	0.14285714286	00641	00601	1
656	601	0.33333333333	00656	00601	1
669	601	0.12500000000	00669	00601	1
698	601	0.16666666667	00698	00601	1
731	601	0.14285714286	00731	00601	1
603	602	0.25000000000	00603	00602	1
610	602	0.14285714286	00610	00602	1
676	602	0.20000000000	00676	00602	1
677	602	0.50000000000	00677	00602	1
602	603	0.25000000000	00602	00603	1
662	603	0.20000000000	00662	00603	1
676	603	0.20000000000	00676	00603	1
690	603	0.50000000000	00690	00603	1

The next step was to calculate total patients in FQHC zip and neighbors. The percentage of patients captured by this method was calculated by dividing the previously calculated total by the total patients from all known zipcodes going to that FQHC. Both these calculations are shown in table 6.

Table 6. Calculating total patients in FQHC zip and neighbors

BHCMISID	FQHC_Zip_Patients	SumOfNeighbor_P	TotNeighborPatients	PatientTot_UnknownC	PercentCaptured
070090	5	4	9	4735	0.19%
051750	36	626	662	183327	0.36%
050030	60	54	114	24095	0.47%
1011880	3	14	17	3051	0.56%
093080	38	101	139	19698	0.71%
024580	22	47	69	9599	0.72%
010040	10	7	17	1211	1.40%
021630	25	895	920	53302	1.73%
106170	2	11	13	658	1.98%
106180	14	11	25	1034	2.42%
033200	306	1643	1949	79107	2.46%
102900	16	47	63	2333	2.70%
040380	100	553	653	23917	2.73%
070530	10	24	34	1219	2.79%
020620	287	261	548	18291	3.00%
021390	1265	251	1516	44787	3.38%
101250	55	53	108	2369	4.56%
020800	158	372	530	11247	4.71%
09E00002	6	102	108	2266	4.77%
080730	314	1677	1991	38838	5.13%

The next step in the EDA process was to examine a non-contiguity based method that systematically eliminated zctas with small numbers of patients. Scenarios for removing 1, 5, 10, 15, 20 and 30 patient zctas were examined both through pivot table analysis as well as spatial techniques. For example, all zctas contributing 1 patient to any FQHC were removed. The total patients remaining after this elimination were aggregated and the percentage of the total FQHC patient population that they represented was calculated. A similar process was executed for the rest of the scenarios.

Two criteria were considered in selecting the optimal threshold for eliminating zctas with small patient populations. The first criterion was the 80% threshold— the point at which most FQHCs are still represented by 80 % of their total population after eliminating small patient zctas. The second criterion was to look at the number of zctas that would be eliminated from the analysis as they consistently contributed small patient populations across all FQHCs. This criterion was important because several zctas have memberships in multiple FQHC Service

Areas. The criterion was intended to isolate zips that consistently contribute small number of patients to all of the FQHCs they are associated with. Conversely, a zip that provides small number of patients to one FQHC but many patients to another will remain.

The performance of the contiguity and non-contiguity based methods were compared. The histograms below show the distributions of the percentage of patients captured by each method (fig.4). The contiguity method shows a high variability (undesirable) with a large number of zeroes. This potentially indicates that Service Areas are not spatially continuous on a consistent basis. The non-contiguity method is clearly superior even after eliminating zctas with small patient populations. The distribution is strongly concentrated in 90%-100% region (desirable).

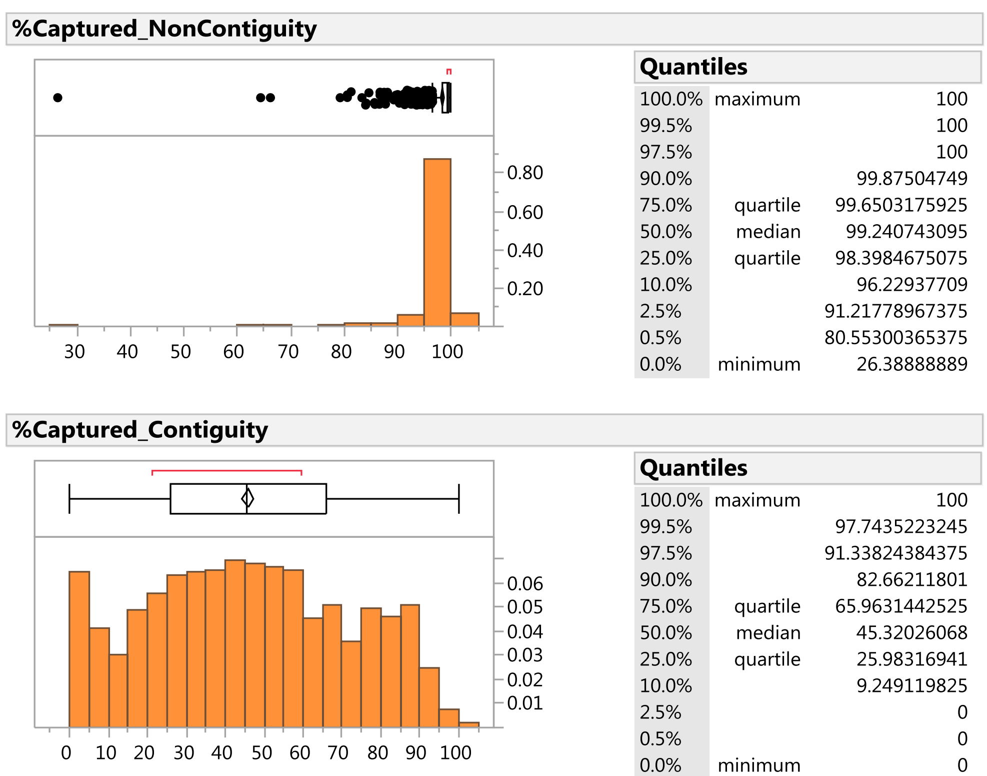


Figure 4. Comparing the contiguity and non-contiguity methods

Different methodologies have been used to resolve the problem with overlapping polygons for service areas. For example, in the case of the HRSA datasets, a plurality rule is applied where a zcta (or other geographical unit) is assigned only to one healthcare provider (hospital, doctor, etc.). When a zcta contributes to multiple service areas, the provider that gets the maximum number of patients retains ownership of the zcta. Accordingly, the zcta is not represented in other service areas. Service area boundaries are further simplified to eliminate overlap and retain contiguity. For example, overlapping service areas may be dissolved so that one service area contains multiple providers. This can also be characterized as more of a patient-centered approach rather a provider-centered one. Another approach is to define a service area based on the plurality of services consumed by the majority population (majority and plurality rules). These methods, however, are more applicable to datasets that contain healthcare consumption data for individuals (Klauss et al, 2005; Goodman et al, 2003; Makuc et al, 1991).

Another simpler method is to limit the service area so that it captures between 80% to 90% of the patient population (Cromley and McLafferty, 2012). This method was better suited to this study for two reasons. First, the dataset does not contain information on individuals. Only aggregate counts for each zipcode are provided. Second, specific communities for each FQHC needed to be identified for assessing performance with respect to cancer screening. This aligns better with ACS goals of improving screening rates within each FQHC, particularly in underperforming ones. After carefully examining the scenarios shown in table 7 and fig.5, the decision was made to eliminate zctas that contributed upto 15 patients (Core Service Areas). This threshold still allowed for upto 99% of the FQHCs to account for 80% of their patient population. Python scripts were revised to eliminate the 15 patient (and less) zctas from each FQHC and recreate new polygons that represented the Core Service Areas shown in fig.6.

Table 7. Scenarios eliminating unique zctas contributing small numbers of patients

Scenario	Number	% of Total ZCTAs
Unique 1 Patient ZCTAs	1720	6.47
Unique 5 Patient ZCTAs	3003	11.29
Unique 10 Patient ZCTAs	3805	14.31
Unique 15 Patient ZCTAs	5629	21.17
Unique 20 Patient ZCTAs	7028	26.43
Unique 30 Patient ZCTAs	9129	34.33

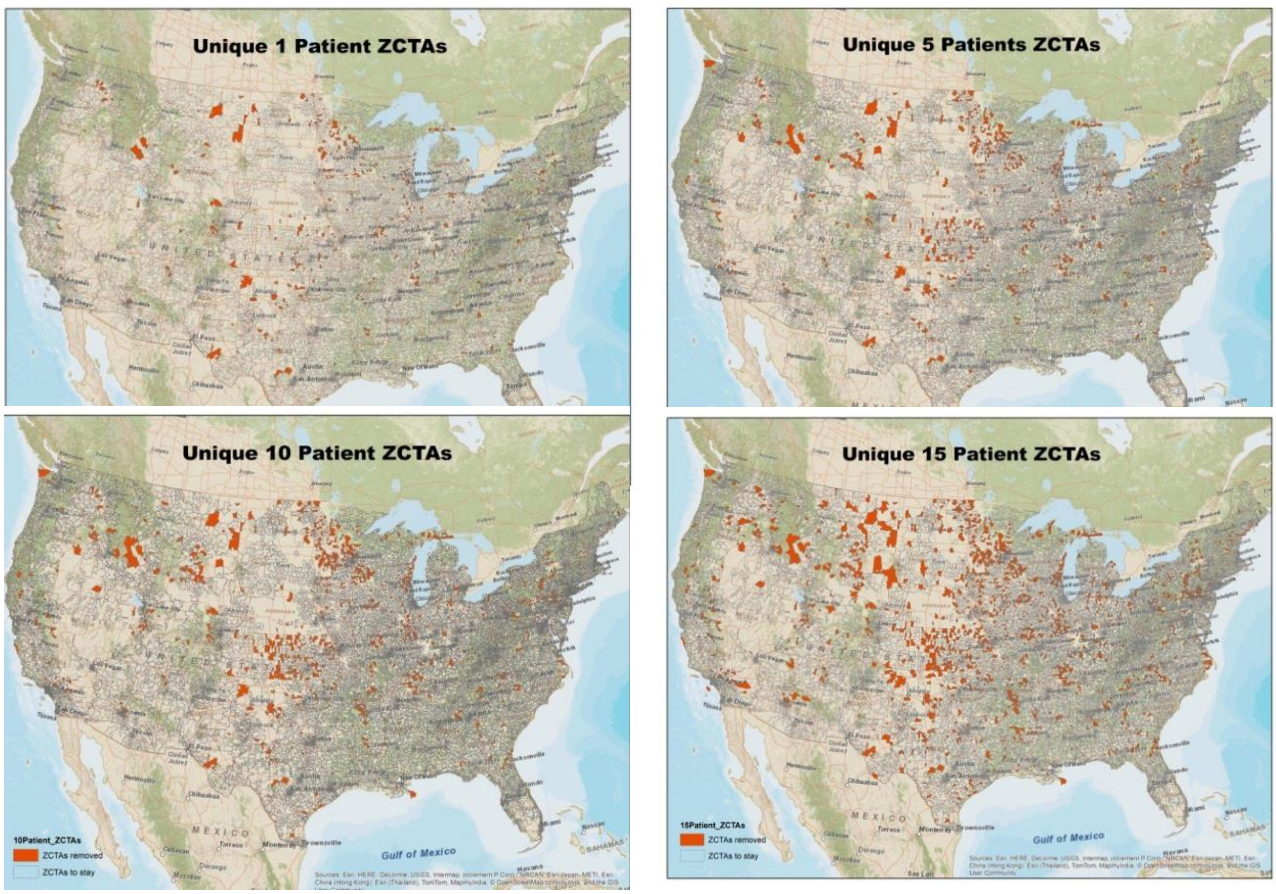


Figure 5. Maps showing the spatial distribution of unique ZCTAs that will be eliminated for each scenario (marked in orange)

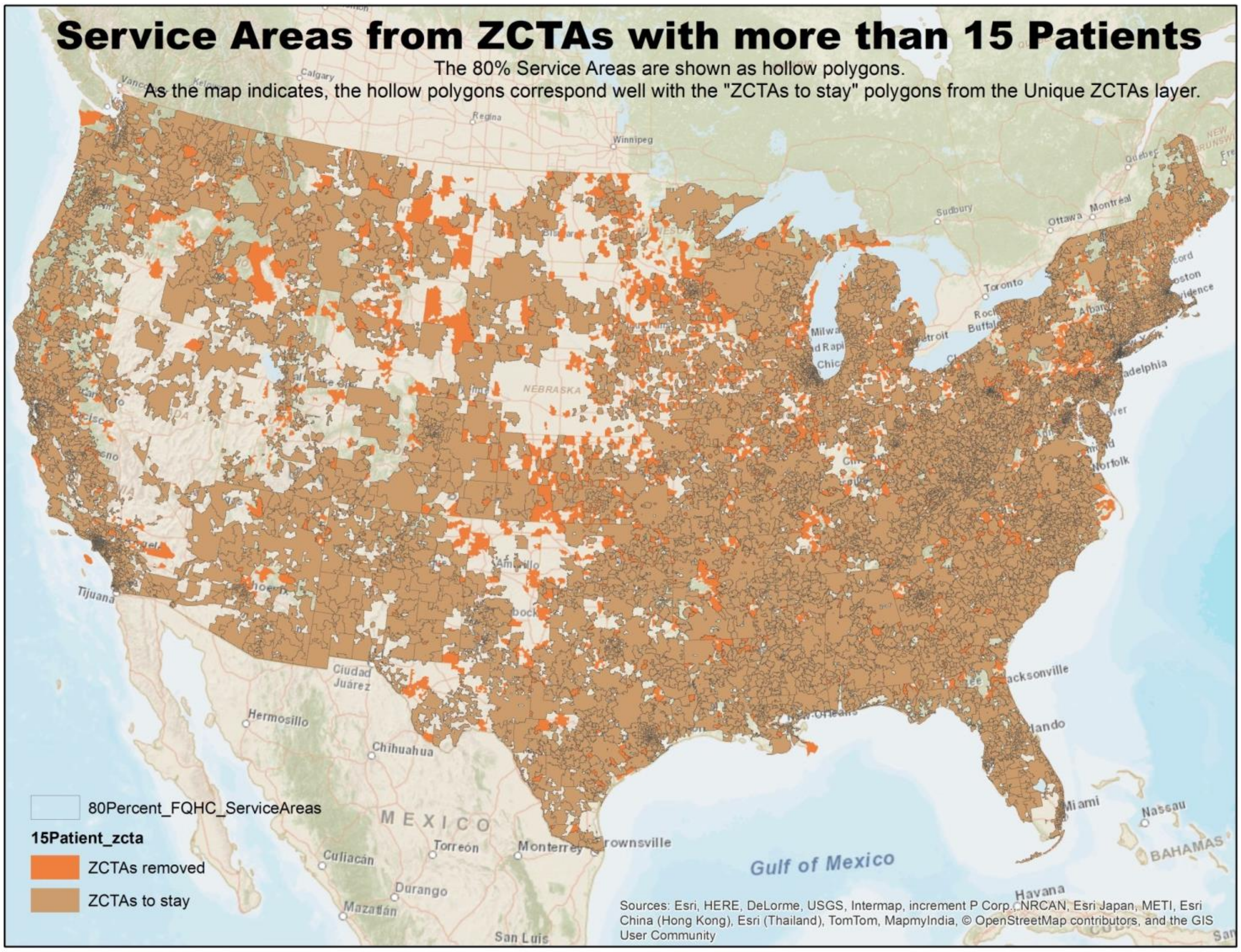
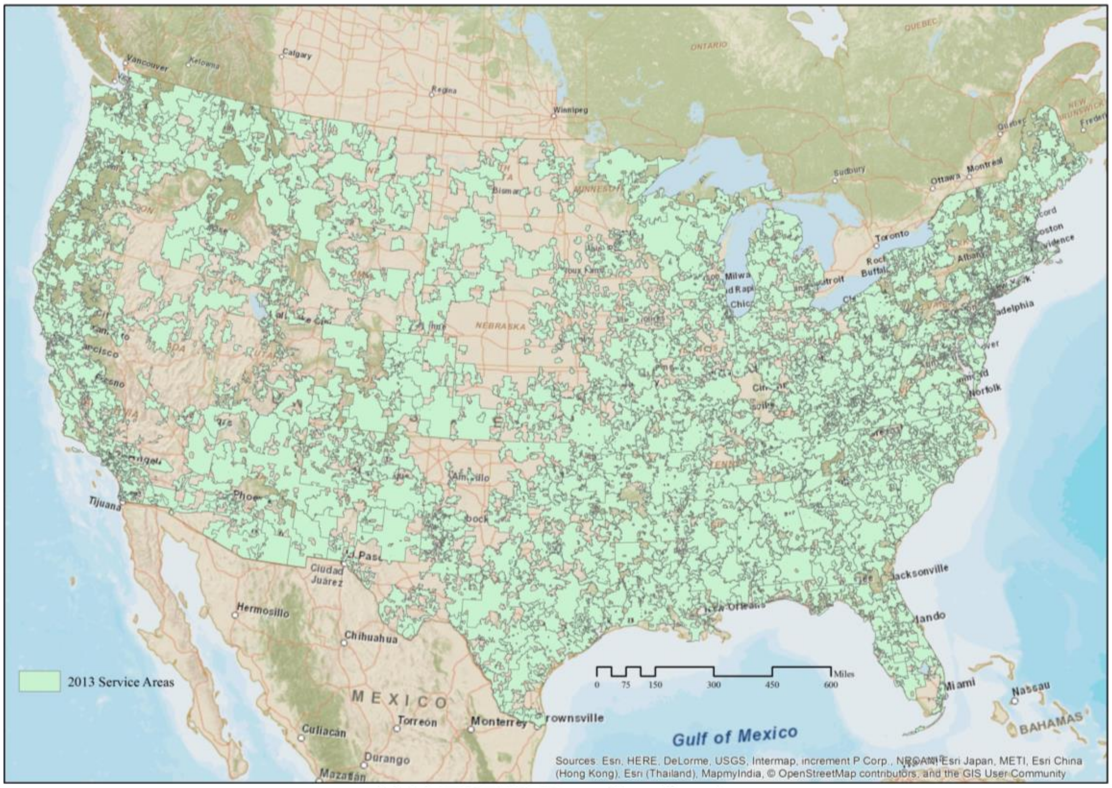


Figure 6. Spatial coverage of Core Service Areas

Validating the Methodology

Service areas for FQHCs in 2013 and 2014 were generated using the identical methodology outlined in the preceding sections to assess its validity for successive years. A preliminary analysis indicated that the methodology performed consistently for the three years. The maps shown in fig.7 help visualize the spatial coverage of FQHC service areas in 2013 and 2014 respectively. There appears to be a marginal improvement in coverage in the lower part of the Midwest (potentially new FQHCs) and will be looked at further in future analysis. The histograms shown in fig.8 help communicate the changes in patient capture from year to year.

2013 FQHC Core Service Areas



2014 FQHC Core Service Areas

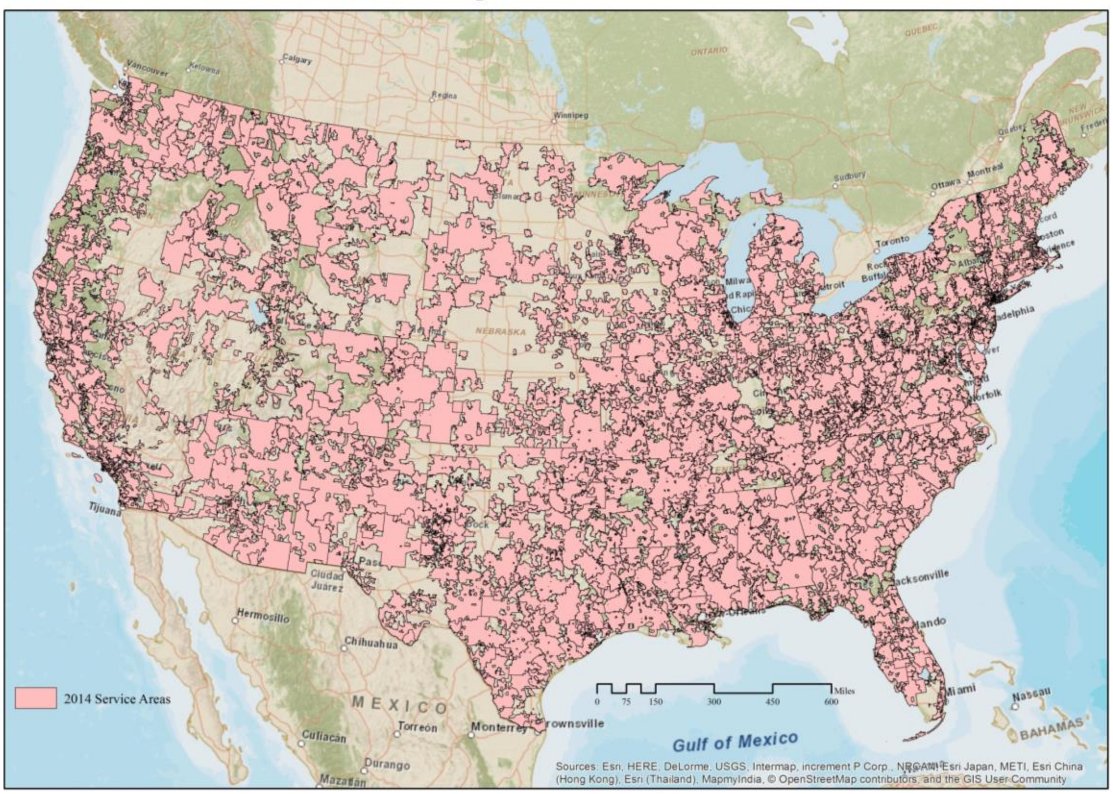
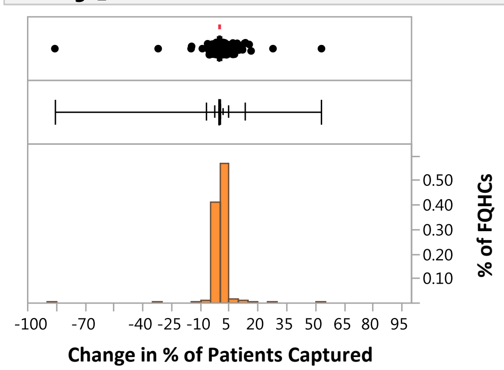


Figure 7. FQHC Core Service Areas for 2013 and 2014.

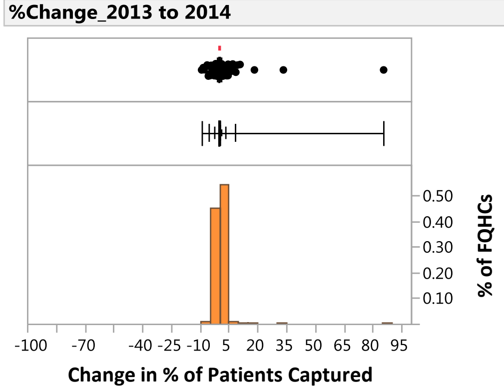
A total of 1167 FQHCs had consecutive data for the years 2012, 2013 and 2014. Percent change in the number of patients captured by the Service Area methodology was compared. There were minor variations between years. Overall, between 2012 and 2014, 719 FQHCs (61%) captured the same or more patients (greater than or equal to 80% of total patient population) based on the methodology developed. Of the 447 FQHCs that saw a decrease, 413 had a decrease of less than 2% in the percentage of patients captured.

%Change_2012 to 2013



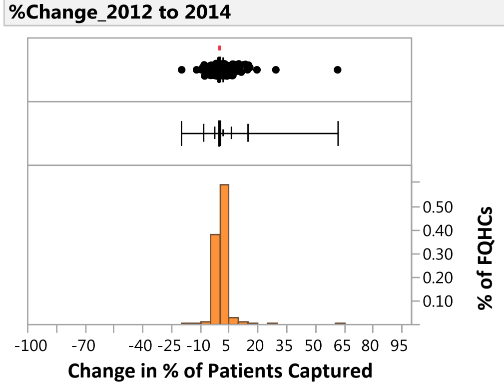
Quantiles		
100.0%	maximum	53.06
99.5%		13.0018
97.5%		4.5385
90.0%		1.524
75.0%	quartile	0.445
50.0%	median	0.04
25.0%	quartile	-0.16
10.0%		-0.634
2.5%		-2.365
0.5%		-6.9433
0.0%	minimum	-85.84

%Change_2013 to 2014



Quantiles		
100.0%	maximum	85.45
99.5%		8.5813000000001
97.5%		3.042
90.0%		0.804
75.0%	quartile	0.26
50.0%	median	0
25.0%	quartile	-0.19
10.0%		-0.7
2.5%		-2.271
0.5%		-5.6525
0.0%	minimum	-9.38

%Change_2012 to 2014



Quantiles		
100.0%	maximum	61.48
99.5%		14.63415
97.5%		6.05775
90.0%		1.696
75.0%	quartile	0.49
50.0%	median	0.07
25.0%	quartile	-0.17
10.0%		-0.62
2.5%		-2.32775
0.5%		-8.07795
0.0%	minimum	-19.77

Figure 8. Trends in percentage of patients captured between 2012, 2013 and 2014.

Identifying Communities of Focus (Phase II)

The next phase in the analysis was to identify areas of geographic focus for ACS interventions. These focus areas represent the cumulative occurrence of risk factors, social vulnerability and low cancer screening rates. Fig.9 outlines the methodological framework for identifying these communities of focus. The framework is focused on colorectal cancer screening due to its significance to ACS initiatives such as the 80% by 2018 NCCRT (national roundtable).

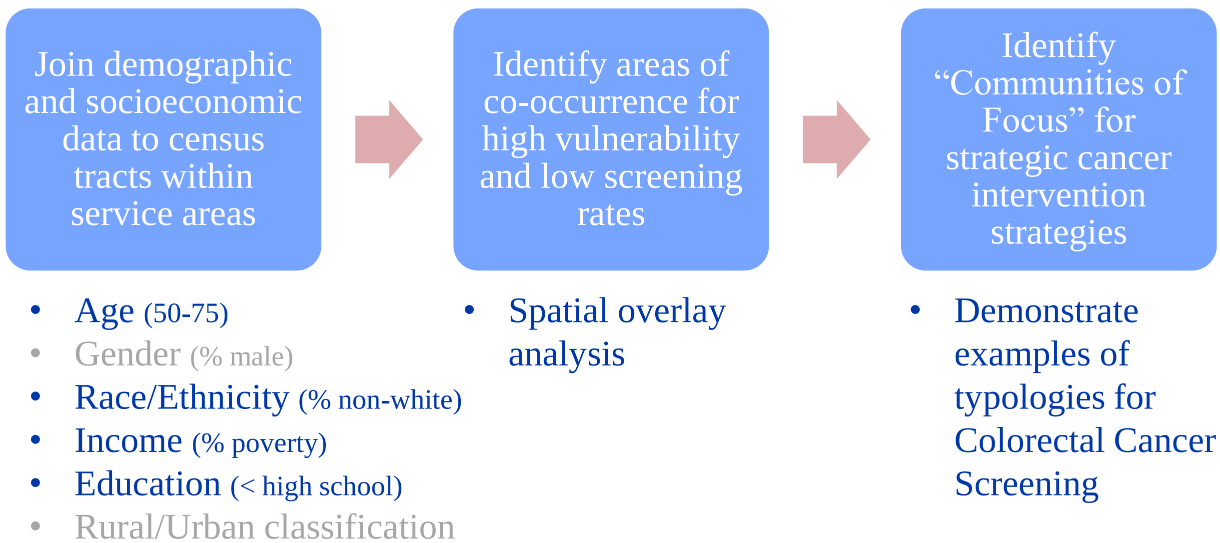


Figure 9. Methodological framework for identifying Communities of Focus

The risk factors (socioeconomic and demographic) shown in the framework were identified directly from prior peer-reviewed research published by ACS (Siegel et al, 2015). The variables in blue were used for immediate analysis. The variables greyed out will be considered in future analysis. Census data for the socioeconomic and demographic variables selected are available at census tract and county scales. The service areas were evaluated against both sets of political boundaries to gauge the appropriate level of analysis.

Counties, particularly in urban areas, have non-uniform population distributions. This is reflected in the size of the tracts. They get smaller towards the urban core, reflecting a higher population density. When service areas only partially cover spatial units, proportioning the population based on area in the tract would be a better approximation rather than area of county included (fig.10). Therefore, census tract level overlay analysis was considered the best suited for the purposes of this project.

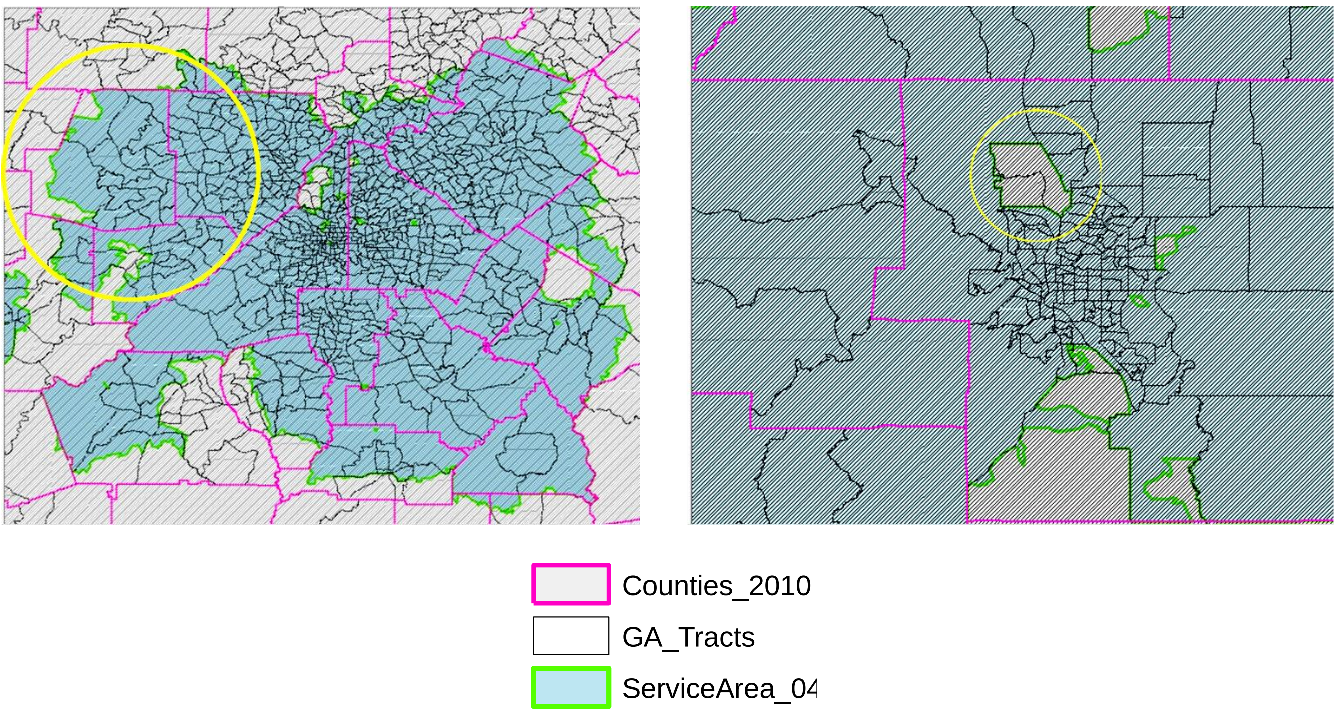


Figure 10. Overlaying counties, census tracts and Service Areas

Based on the algorithm illustrated in fig.11, python scripts were written to intersect each service area with census tract boundaries. Areas of the tracts within each Service Area were calculated. Tracts which were partially included within the service area will have reduced areas (IntSubArea) compared to their original areas before they were intersected (Area1). The percentage area of the tract represented within the service area was used as a weight to proportion the population and associated socioeconomic and demographic variables (fig.12). The resulting dbf tables from each of the service area shapefiles were appended into a single excel table for further analysis.

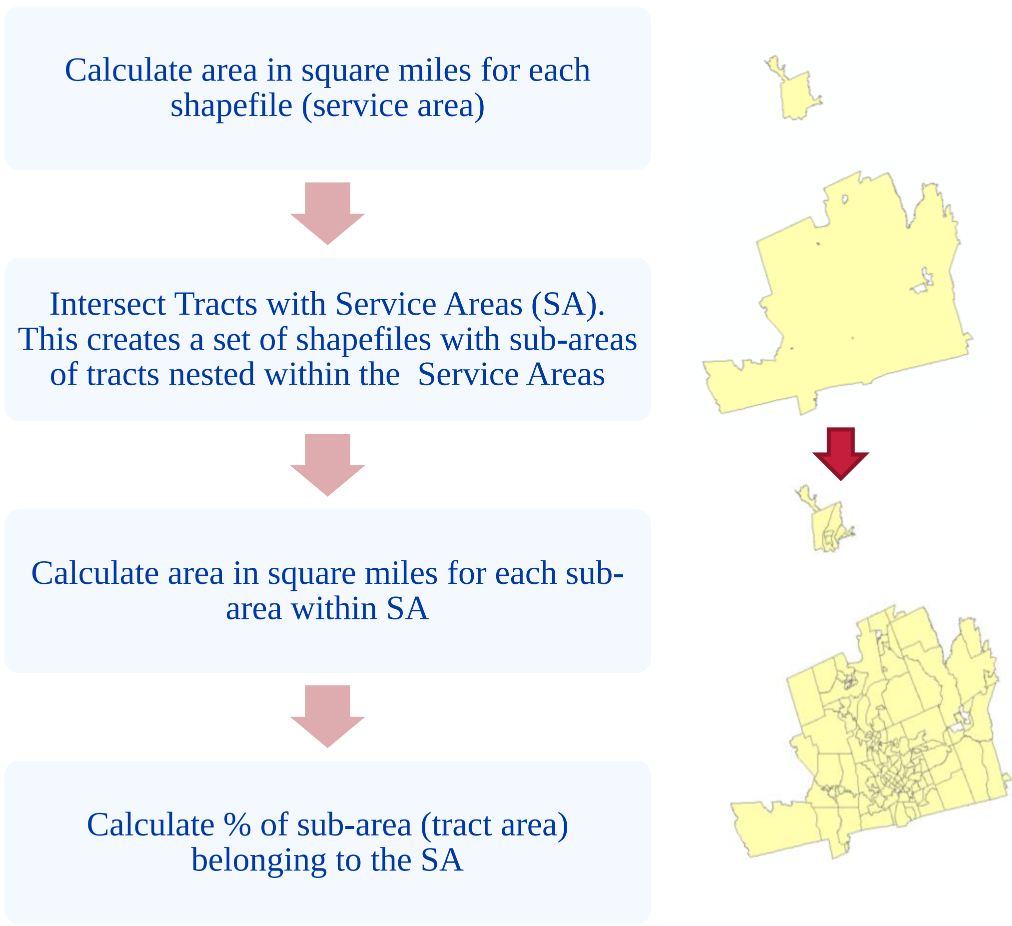


Figure 11. Algorithm for intersecting Service Areas with census tracts

BHCMISID	SA_Area	FID_Tracts	GEOID10	Area1	IntSubArea
010030	825.585826	31736	25011040702	29.090613	0.959684
010030	825.585826	31737	25011040701	2.366821	2.366821
010030	825.585826	31738	25011040600	108.757225	3.119994
010030	825.585826	31739	25011041400	1.737068	1.737068
010030	825.585826	31757	25011041300	0.927188	0.927188
010030	825.585826	31758	25011041200	0.845401	0.845401
010030	825.585826	31759	25011041100	7.315088	7.315088
010030	825.585826	31760	25011041000	11.058449	11.058449
010030	825.585826	31765	25011040200	56.222568	4.143512
010030	825.585826	31785	25015822603	27.356162	27.356162
010030	825.585826	31787	25015821903	1.166175	1.166175
010030	825.585826	31801	25015821904	0.47116	0.47116
010030	825.585826	31802	25015820101	37.143528	37.143528
010030	825.585826	31803	25015820202	26.553697	26.553697
010030	825.585826	31804	25015820204	35.310024	35.310024

FQHC ID and FQHC SA area in sq miles

Tracts that the SA intersects with

Area1 is the area of the full tract and IntSubArea is the area of the portion of that tract within the SA

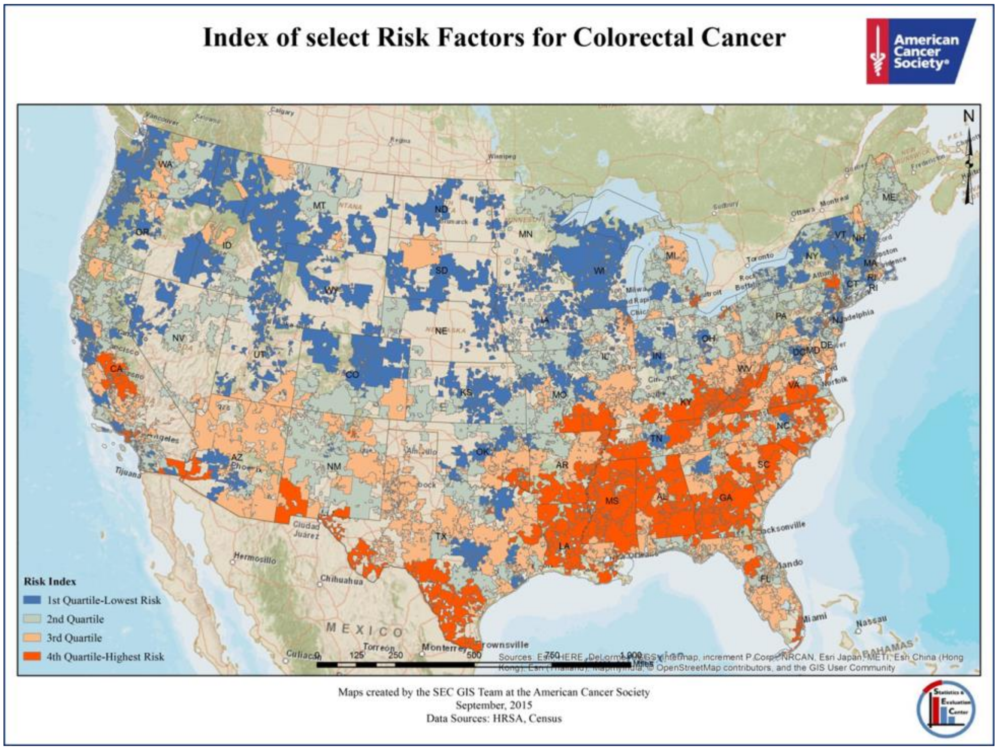
Figure 12. Attribute tables with percentage of tract area calculated

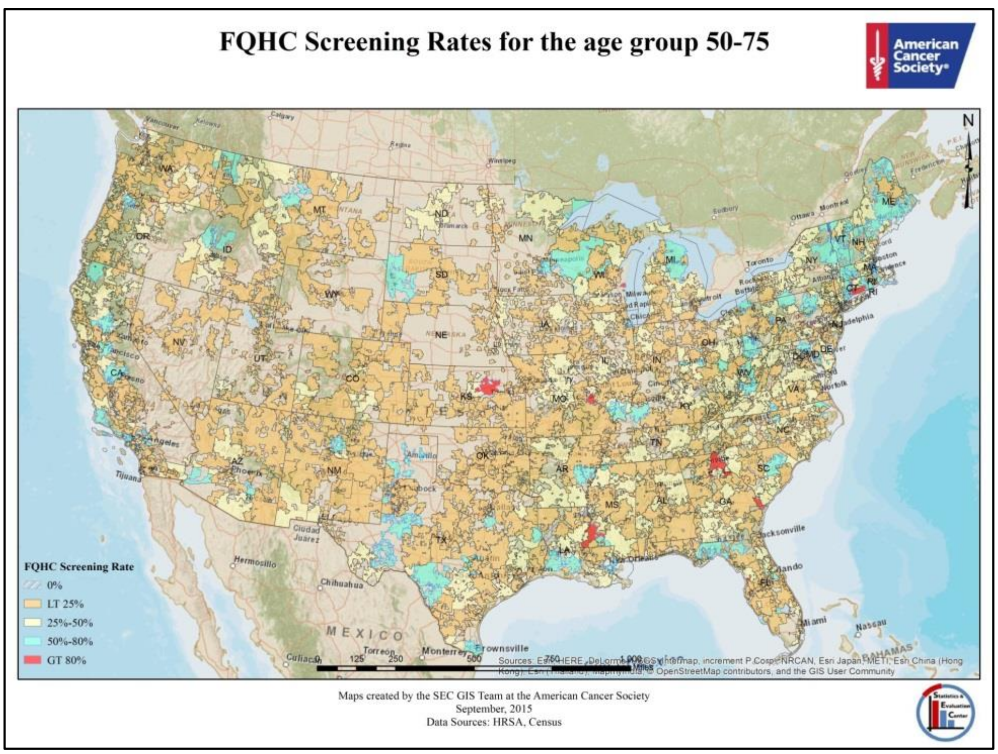
The relevant socioeconomic and demographic variables were downloaded from the American Community Survey. 2012 5- year estimates were used for the analysis. This table served as a master vlookup table for joining with the appended attribute tables. Using the VLOOKUP function in excel, each tract within every Service Area was assigned the matrix of risk factor variables. The variables were multiplied by the respective tract weights (IntSubArea) to calculate the proportion within each Service Area. Pivot tables were used to aggregate the totals for each variable by the FQHC Service Area (table 8). This restructures the data from a stacked table (multiple rows for each FQHC) to a table with one row of variables per FQHC.

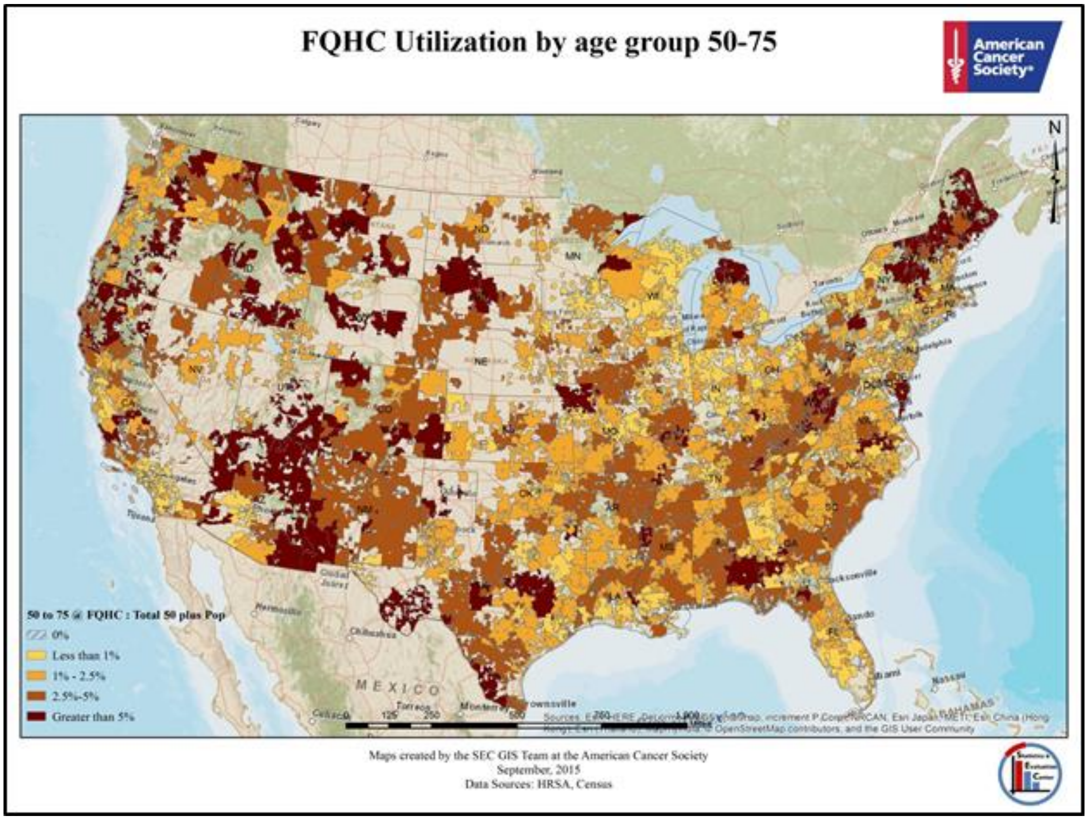
Table 8. Risk factor variables proportioned and aggregated to the service area level

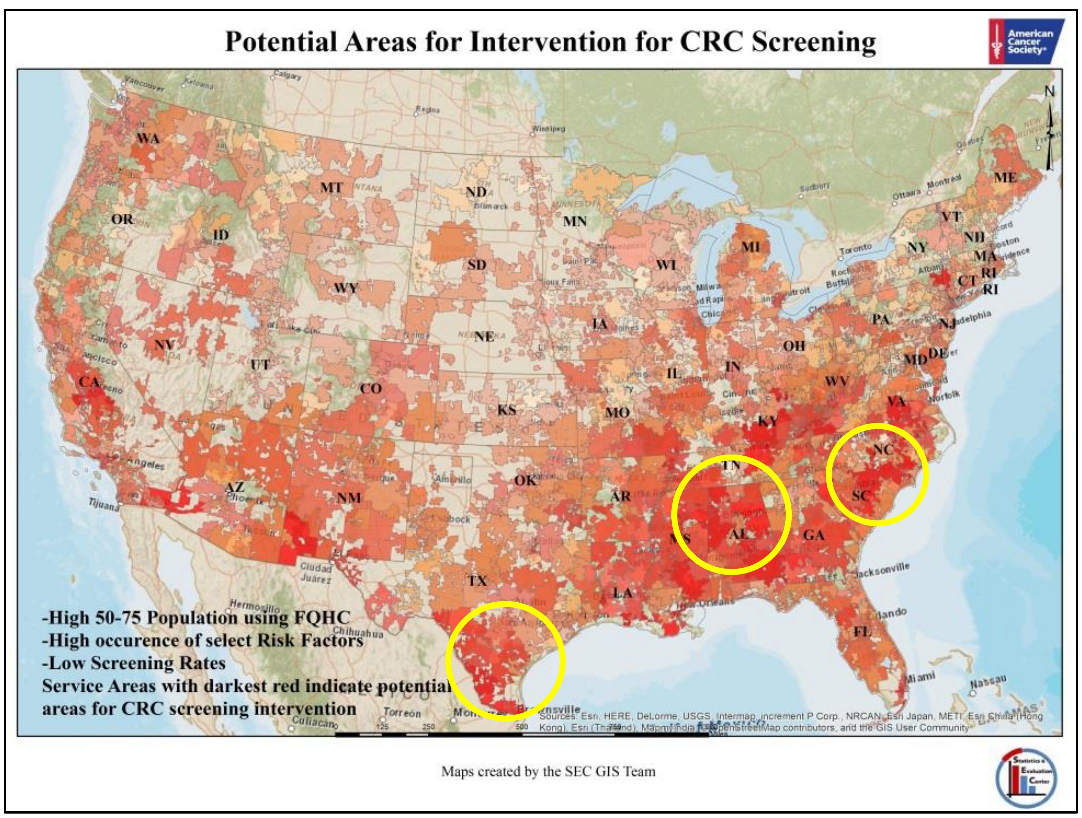
BHCMISID	Tot_Pop	50_plus	Male	M_50_plus	Fem	F_50_plus	White	Black	Poverty	45_65
010030	606729.9352	207684.6777	289062.077	93449.66467	317667.8579	114235.01	481401.1	45161.57	99374.88	16705
010040	90927.34717	33996.91151	44569.1712	15907.59387	46358.17599	18089.318	83898.06	3604.767	15149.34	26458
010060	397479.8577	128365.2106	190341.385	57011.20716	207138.4726	71354.003	265444.2	74306.02	55639.01	10402
010070	1647132.677	541056.2644	797176.338	244105.948	849956.3386	296950.32	1181848	221090.9	209144.3	44770
010120	396801.8232	125411.4181	188500.233	55849.94618	208301.5903	69561.472	286738.6	40089.94	83564.43	99303
010130	99590.62809	29743.36209	49388.8801	13717.59188	50201.74799	16025.77	85107.33	4285.63	14884.37	25018
010150	271653.2335	88430.16935	132790.114	41191.22562	138863.1196	47238.944	253998.4	4548.216	30510.63	76459
010160	850808.9928	217419.878	413795.991	98026.23573	437013.0013	119393.64	574337.2	83252.18	142436.2	17818
010170	1963508.749	641769.6489	944037.723	286262.0104	1019471.026	355507.64	1384759	262712.8	262905	51584
010180	221689.775	63712.48725	107921.216	28770.7703	113768.559	34941.717	157726.9	22985.47	44446.57	52562
010220	491705.5731	165386.8816	244974.003	77755.27583	246731.5704	87631.606	415024.9	25126.07	46035.1	14094
010290	564778.2522	162369.2279	268570.046	70913.44604	296208.2063	91455.782	268126.4	175525.7	108033.8	13803
010330	554141.3746	193842.8414	263219.356	87244.16326	290922.0186	106598.68	449410.7	38297.36	86808.53	15209
010340	141776.3902	54417.77242	69148.0223	25115.39294	72628.36785	29302.379	135998.3	1052.134	21032.31	43694
010380	24936.93717	11121.91129	12270.8547	5281.969466	12666.08252	5839.9418	23513.24	118.0457	4073.812	8000
010450	7933.951882	3297.011339	3722.27135	1416.261881	4211.680536	1880.7495	7613.069	46.14149	1141.568	2465.
010460	460318.1765	176811.8693	224350.497	82454.65528	235967.6792	94357.214	436771.5	7618.522	66921.53	14055
010480	931728.167	320173.0683	450389.892	144820.7059	481338.2753	175352.36	760712.4	52987	121311.1	26237
010530	49096.31209	20931.52513	23890.385	9739.432204	25205.9271	11192.093	46957.07	238.1842	7568.031	15187
010570	26999.37925	11964.80783	13277.4437	5688.666695	13721.93556	6276.1411	25173.51	123.471	4556.268	8610.
010580	843656.3353	283229.798	405819.021	126507.8692	437837.3144	156721.93	659732.4	57374.85	121505	22719
010600	3267139.287	1022660.298	1572299.68	457920.2505	1694839.607	564740.05	2386612	316517	395824.3	84980

Percentages for each variable were calculated based on the total population in the service areas. Percentages make the service areas comparable. The percentages were further normalized by calculating z-scores and added up to calculate a composite index score of risk factors. The risk index was overlaid with the screening rates and with FQHC utilization rates by the 50-75 age group. The layers were overlaid using transparent color fill (graduated red tones) for the polygons (service areas). When the layers were placed over each other, it became easy to visualize “Communities of Focus” with the darkest color (highest intensity of red). The series of maps below show the individual layers as well as the results of the overlay analysis.


Index of select Risk Factors for Colorectal Cancer
American Cancer Society
Risk Index
1st Quartile-Lowest Risk
2nd Quartile
3rd Quartile
4th Quartile-Highest Risk
Maps created by the SEC GIS Team at the American Cancer Society
September, 2015
Data Sources: HRSA, Census


FQHC Screening Rates for the age group 50-75
American Cancer Society
FQHC Screening Rate
0%
LT 25%
25%-50%
50%-80%
GT 80%
Maps created by the SEC GIS Team at the American Cancer Society
September, 2015
Data Sources: HRSA, Census


FQHC Utilization by age group 50-75
American Cancer Society
50 to 75 @ FQHC : Total 50 plus Pop
0%
Less than 1%
1% - 2.5%
2.5%-5%
Greater than 5%
Maps created by the SEC GIS Team at the American Cancer Society
September, 2015
Data Sources: HRSA, Census


Potential Areas for Intervention for CRC Screening
American Cancer Society
-High 50-75 Population using FQHC
-High occurence of select Risk Factors
-Low Screening Rates
Service Areas with darkest red indicate potential
areas for CRC screening intervention
Maps created by the SEC GIS Team

Contributions to ACS initiatives

This analysis is envisioned to contribute to the following ACS initiatives:

- **National Colorectal Cancer Roundtable- 80% x 2018**
  - Identify FQHC for focused intervention for CRC screening
- **Cervical Cancer Prevention**
  - As part of the HPV Roundtable effort, identify FQHCs to promote HPV vaccination efforts (immunization)
  - Identify FQHCs for increasing utilization of pap tests (screening)
- **Lung cancer prevention**
  - Identify FQHCs for increased efforts towards Tobacco use assessment and intervention
  - **Breast Cancer prevention**
  - Identify areas of focus for increasing breast cancer prevention efforts (mammograms, exams)

Directions for Further Research

Preliminary statistical analysis did not yield significant results. While the methodology was useful in identifying vulnerable communities, the determinants of screening rates might be internal to the FQHC. Therefore this approach might not be as effective in predicting screening rates based on community-level characteristics. More in-depth analysis will be pursued to better understand mechanisms internal t the FQHC that are more closely related to screening rates.

Meanwhile, methods such as hierarchical clustering will be applied to identify and categorize these communities of focus to inform the design of strategic interventions. Other techniques of index construction will also be explored.

The particular requirement for reporting screening rates to the UDS system was initiated in 2012. It is possible that there may be errors associated with data reporting as FQHCs tried to cope with the new reporting protocols. The analysis carried out on 2012 data will be extended to 2013 and 2014 data to examine data consistency and validity. The analytical approach will also be applied to other cancer prevention initiatives outlined in the previous section.

References:

Allen, C. L., Harris, J. R., Hannon, P. A., Parrish, A. T., Hammerback, K., Craft, J., & Gray, B. (2013). Opportunities for Improving Cancer Prevention at Federally Qualified Health Centers. Journal of Cancer Education, 29(1), 30-37. doi:10.1007/s13187-013-0535-4

Cromley, E.K. and McLafferty, S.L. (2012). GIS and Public Health (2nd edition). Guilford Press, New York, NY.

Daly, J., Levy, B., Moss, C., & Bay, C. (2015). System Strategies for Colorectal Cancer Screening at Federally Qualified Health Centers. American Journal of Public Health, 105(1), 212-219.

Goebel, J. (2013). A Brief History of Federally Qualified Health Centers (FQHC). http://www.notifymd.com/a-brief-history-of-federally-qualified-health-centers-fqhc/.

Goodman, D., Mick, S., Bott, D., Stukel, T., Chang, C., Marth, N., . . . Carretta, H. (2003). Primary Care Service Areas: A New Tool for the Evaluation of Primary Care Services. Health Services Research, 38(1p1), 287-309.

Klauss, G., Staub, L., Widmer, M., & Busato, A. (2005). Hospital service areas – a new tool for health care planning in Switzerland. BMC Health Services Research, 5, 33.

Makuc DM et al. (1991). Health service areas for the United States National Center for Health Statistics. Vital Health Stat (2)1 12, U.S. Department of Health and Human Services.

Martinez-Gutierrez, J., Jhingan, E., Angulo, A., Jimenez, R., Thompson, B., & Coronado, G. (2013). Cancer Screening at a Federally Qualified Health Center: A Qualitative Study on Organizational Challenges in the Era of the Patient-Centered Medical Home. Journal of Immigrant and Minority Health, 15(5), 993-1000. doi:10.1007/s10903-012-9701-8

Taplin, S. H., Haggstrom, D., Jacobs, T., Determan, A., Granger, J., Montalvo, W., . . . Calvo, A. (2008). Implementing Colorectal Cancer Screening in Community Health Centers: Addressing Cancer Health Disparities through a Regional Cancer Collaborative. Medical Care, 46(9), S74-S83.

Taylor, J. (2012). Changes in Latitudes, Changes in Attitudes: FQHCs and Community Clinics in a Reformed Health Care Market.

Siegel RL, Sahar L, Robbins A, Jemal A.(2015). Where can colorectal cancer screening interventions have the most impact? Cancer Epidemiol Biomarkers Prev. 24(8):1151-6. doi: 10.1158/1055-9965.EPI-15-0082. Epub 2015 Jul 8.